

SEGMENTASI BAGIAN TUBUH MANUSIA 3D *POINT CLOUD* BERBASIS *SPHERICAL PROJECTION* MENGGUNAKAN *MASK R-CNN*

Muhammad Alwi Dahlan^{1*}, Aziz Musthafa¹, Oddy Virgantara Putra¹

¹Teknik Informatika, Universitas Darussalam Gontor, Ponorogo

Jl. Raya Siman, Dusun I, Demangan, Kec. Siman, Kabupaten Ponorogo, Jawa Timur

*Corresponding Author : muh.alwidahlan19@mhs.unida.gontor.ac.id

Abstrak

Teknologi sensor yang berkembang pesat, yaitu Light Detection and Ranging (LiDAR), banyak digunakan dalam bidang segmentasi citra dan deteksi objek. Point cloud 3D, yang memvisualisasikan pantulan objek, menjadi hasil dari teknologi ini. Namun, kendala atau tantangan muncul dalam segmentasi point cloud 3D karena struktur data yang tidak beraturan, menyebabkan noise dan tumpang tindih bentuk objek. Hal ini memengaruhi keakuratan segmentasi dan deteksi objek. Untuk mengatasi masalah tersebut, jawabannya ditemukan dalam pendekatan proyeksi bola. Dengan mengubah data point cloud 3D menjadi gambar 2D, proses segmentasi gambar menjadi lebih mudah. Tujuannya adalah untuk segmentasi bagian tubuh dengan menggunakan data 3D point cloud, Saat melakukan segmentasi gambar hasil konversi data 3D ke 2D menggunakan deep learning, model yang efektif untuk mempelajari data tersebut adalah Mask R-CNN. Akurasi 59% berhasil dicapai dalam 200 epoch berdasarkan data pelatihan. Penelitian ini diharapkan dapat membuka jalan untuk penelitian lebih lanjut guna menyempurnakan dan mengembangkan ilmu pengetahuan dibidang computer vision dan kecerdasan buatan.

Kata kunci: *LiDAR; Point Cloud; Segmentasi; Spherical Projection*

Abstract

The rapidly developing sensor technology Light Detection and Ranging (LiDAR) is widely used in the fields of image segmentation and object detection. 3D point clouds, which visualize object reflections, are the result of this technology. However, obstacles or challenges arise in 3D point cloud segmentation due to the irregular structure of the data, causing noise and overlapping of object shapes. This affects the accuracy of segmentation and object detection. To overcome such problems, the answer is found in the spherical projection approach. By converting 3D point cloud data into 2D images, the image segmentation process becomes easier. The goal is to segment body parts using 3D point cloud data, When segmenting images converted from 3D to 2D data using deep learning, an effective model for learning such data is Mask R-CNN. An accuracy of 59% was achieved in 200 epochs based on the training data. This research is expected to pave the way for further research to enhance and develop science in the field of computer vision and artificial intelligence.

Keywords : *LiDAR; Point Cloud; Segmentasi; Spherical Projection*

PENDAHULUAN

Segmentasi bagian tubuh manusia merupakan masalah mendasar dalam *computer vision* dan memiliki berbagai aplikasi seperti interaksi manusia-komputer, *virtual reality*, *augmented reality*, dan robotika. Segmentasi bagian tubuh secara akurat, gerakan manusia

dapat dianalisis, gerakan tangan dilacak, *pose* dikenali, dan interaksi virtual realistis.

Segmentasi merupakan salah satu elemen terpenting dalam analisis citra otomatis, terutama karena pada fase ini objek atau entitas lain yang diinginkan diekstraksi dari citra untuk diproses lebih lanjut, seperti deskripsi dan

identifikasi. Misalnya, citra satelit yang menunjukkan area darat dan laut harus mensegmentasi objek di sisi darat dengan benar agar dapat diklasifikasikan dengan lebih akurat (Marpaung et al. 2022) Masalah yang mengacu pada tugas mengidentifikasi dan membedakan berbagai bagian tubuh manusia secara akurat menggunakan representasi *point cloud*. Masalah ini biasa terjadi pada visi komputer dan rekonstruksi 3D yang tujuannya adalah untuk menganalisis dan memahami gerakan, interaksi, atau struktur anatomi manusia.

Point cloud adalah kumpulan titik 3D yang ditangkap oleh berbagai sensor, seperti kamera kedalaman atau *LiDAR* (*Light Detection and Ranging*), yang merupakan teknologi penginderaan jauh aktif yang dapat memberikan awan titik pada dari objek yang dipindai (Liu et al. 2020) Segmentasi bagian tubuh manusia dalam *point cloud* sulit dilakukan karena sifat anatomi manusia yang kompleks, variasi *pose* dan penampilan. Mengidentifikasi dan membedakan awan titik manusia secara akurat seringkali sulit, terutama dalam situasi dimana ditutupi oleh sesuatu dan tidak memiliki sinyal warna atau tekstur (Ryselis et al. 2022)

Mask R-CNN (*Region-based Convolutional Neural Network*) adalah arsitektur *deep learning* yang menggabungkan deteksi objek dan segmentasi instan. Bellynya and Syaputra (2022) menjelaskan bahwa *Mask R-CNN* diperkenalkan pada tahun 2017 oleh tim *Facebook AI Research (FAIR)* yang merupakan kerangka *deep learning* yang mengenali objek dalam gambar dan membuat segmentasi mask untuk setiap objeknya. Masalah spesifik segmentasi bagian tubuh manusia di *point cloud* menggunakan *Mask R-CNN* merupakan melatih jaringan saraf untuk mengenali dan membedakan bagian tubuh individu seperti kepala, badan, dan lengan, di *cloud* titik tertentu. Jaringan dilatih menggunakan kumpulan data berlabel, dengan setiap titik ditetapkan ke kelas bagian tubuh tertentu. Jaringan belajar untuk mengekstrak fitur yang relevan dari data *point cloud* dan memprediksi yang akurat setiap bagian tubuh.

Adanya kemajuan teknologi dan komputer memungkinkan mengidentifikasi bagian tubuh manusia dengan memanfaatkan *artificial intelligence* dengan menggunakan data *3D point cloud* yang dihasilkan

menggunakan sensor *LiDAR* dengan menggunakan algoritma *Mask R-CNN*.

PENELITIAN TERDAHULU

Untuk penelitian ini, peneliti membuat perbandingan temuan penelitian ini dengan penelitian terdahulu:

Unknown 3D Objects From Real Depth Images

Penelitian Danielczuk et al. (2019) berjudul “*Segmenting Unknown 3D Objects From Real Depth Images Using Mask R-CNN Trained On Synthetic Data*” membahas kemampuan untuk membedakan objek yang tidak diketahui karena segmentasi akurat benda yang tidak diketahui di lingkungan yang tidak beraturan untuk meningkatkan kemampuan robot dalam menangkap dan melacak. Masalah pada penelitian ini adalah tantangan segmentasi objek 3D yang tidak diketahui dari citra kedalaman nyata dalam lingkungan yang penuh dengan berbagai objek (Krähenbühl and Koltun 2014), tanpa memerlukan data yang diberi label secara manual (T.-Y. Lin et al. 2014). Penelitian ini membahas untuk mengembangkan metode yang dapat secara efektif mensegmentasi objek di lingkungan yang beragam dan juga keterbatasan metode segmentasi yang sudah ada dengan mengeksplorasi data sintesis dan randomisasi domain untuk meningkatkan kinerja segmentasi.

Dengan metode *Synthetic Depth Mask R-CNN*, metode segmentasi versi baru yang menggabungkan segmentasi berbasis *point cloud* dan segmentasi berbasis *deep learning*. *SD Mask R-CNN* yang mewakili kemampuan generalisasi yang baik dari data simulasi di dunia nyata dan lingkungan simulasi. Jaringan ini mencapai presisi dan *recall* rata-rata pada kumpulan data beresolusi tinggi dan menunjukkan kemampuan generalisasi ke sensor beresolusi rendah. Metode *SD mask R-CNN* digunakan dalam penelitian ini dan menunjukkan segmentasi objek yang efektif dengan tingkat keberhasilan (74%) melampaui pengklasifikasi dasar *Euclidean PCL* (56%) dan teknik *Mask R-CNN* yang disesuaikan pada gambar berwarna nyata (78%). Selain itu, penelitian ini mengevaluasi metode *SD Mask R-CNN* menunjukkan peningkatan signifikan dalam presisi rata-rata dan perolehan dibandingkan dengan metode segmentasi berbasis *cluster* dari *PCL* (Hochman et al.

1977). Hasil ini menunjukkan bahwa metode *SD mask R-CNN* tahan terhadap gangguan sensor dan mengungguli metode segmentasi lainnya pada berbagai kumpulan data pengujian.

Hasil penelitian penelitian ini fokus pada pengembangan metode segmentasi objek menggunakan data sintetik dan teknik *deep learning*. Hasil eksperimen menunjukkan bahwa meningkatkan ukuran kumpulan data pelatihan sintetik dan jumlah objek unik yang digunakan untuk pelatihan akan meningkatkan performa model segmentasi. Studi ini menunjukkan bahwa kinerja dapat lebih ditingkatkan dengan menambahkan lebih banyak data pelatihan.

Improve In Edge Segmentation

Penelitian Hong Hai and Bao Long (2021) berjudul “*Improve Of Mask R-CNN In Edge Segmentation*” membahas keterbatasan model *Mask R-CNN* dalam mengidentifikasi objek secara akurat dilokasi yang sulit terutama di tepi dan batas. Penulis menunjukkan bahwa *Masks* asli *Mask R-CNN* seringkali tidak menutupi seluruh objek, menyebabkan kesulitan dalam penentuan posisi objek dan perencanaan tindakan. Penelitian ini berfokus pada peningkatan kualitas segmentasi objek terutama di tepi. Menggabungkan *Mask R-CNN* dengan algoritma 3D untuk mengklasifikasikan wilayah tepi secara akurat (Rao et al. 2010).

Dengan mengembangkan model *Mask R-CNN* yang dapat deteksi objek baru dengan menambahkan cabang prediksi *Mask*. *Mask* ini digunakan untuk menutupi dan memisahkan area objek. Model ini mampu membuat mask yang sangat baik yang menutupi hampir seluruh target bahkan dalam situasi multi-kelas. Dengan metode inovatif yang menggabungkan alur kerja *Mask R-CNN* asli dan cabang algoritma 3D untuk mempertahankan dan mengklasifikasi wilayah *edge* dengan lebih baik. Hasil dari penelitian ini menyatakan bahwa peningkatan kinerja *Mask R-CNN* dengan segmentasi objek yang lebih detail (Wu, Wen, and Xie 2019). Dalam situasi sulit, peningkatan signifikan dalam menemukan objek tertentu.

Hasil dari penelitian ini menunjukan bahwa metode yang diusulkan dengan menggabungkan *Mask R-CNN* dengan algoritma 3D untuk mensegmentasi objek dalam cakupan robot *grasping* (Uckermann et

al. 2012), meningkatkan kinerja detail segmentasi. Para penulis menunjukan bahwa penambahan segmentasi *Difference of Normals (DoN)* dan Ekstraksi *Klaster Euclidean* membantu mempertahankan dan mengklasifikasi daerah tepian, mengatasi keterbatasan *Mask R-CNN* untuk mengidentifikasi objek secara akurat diketerbatasan *Mask R-CNN* untuk mengidentifikasi objek secara akurat ditempat yang sulit. Hasil percobaan menunjukkan bahwa metode yang diusulkan mencapai hasil segmentasi *Mask R-CNN*.

Segmentation And Tracking For MADS Dataset

Penelitian Le and Scherer (2021) berjudul “*Human Segmentation And Tracking Survey On Masks For MADS Dataset*” membahas tentang kumpulan data *MADS (Martial Arts, Dancing and Sports)* (Zhang et al. 2019), yang mencakup aktifitas kompleks seperti seni bela diri dan tarian, menimbulkan tantangan dalam melakukan segmentasi dan pelacakan manusia dalam gambar dan video menggunakan *Convolutional Neural Network (CNN)* (Xu et al. 2019). Segmentasi manusia adalah pemisahan data manusia dari data lain dalam adegan kompleks dan penting untuk diidentifikasi dan pemantauan aktifitas. Pelacakan yang berfokus pada penggalian lokasi seseorang dari video dan diterapkan dalam pekerjaan seperti pengamatan dan pengawasan (Yao et al. 2019). Penelitian ini membahas tantangan yang terkait dengan mengidentifikasi dan melacak manusia dengan akurasi dalam data visual, yang merupakan masalah penting dalam *computer vision*.

Penelitian ini menggunakan berbagai *deep learning*, khususnya *Convolutional Neural Network (CNN)*, untuk mensegmentasi dan melacak orang dalam gambar dan video. Dengan metode yang digunakan *Convolutional Neural Networks (CNN)* untuk implementasi seperti *Fast R-CNN*, *Faster R-CNN*, *YOLO*, *AlexNet*, dan *VGGNet* yang digunakan untuk tugas deteksi dan segmentasi objek, termasuk segmentasi manusia. Memanfaatkan *deep learning* untuk secara akurat mengidentifikasi data manusia dari adegan yang kompleks. Mengevaluasi hasil segmentasi manusia menggunakan dataset *MASK MADS* dan melaporkan metrik kinerja seperti *Average*

Precision (AP) untuk segmentasi manusia dan pelacakan manusia.

Hasil penelitian dan eksperimen penelitian ini fokus pada segmentasi dan pelacakan manusia menggunakan teknik *deep learning* khususnya *Convolutional Neural Network (CNN)*. *Convolutional Neural Networks (CNN)* telah secara signifikan meningkatkan kinerja pelacakan dan segmentasi manusia dalam gambar dan video. Penggunaan model *CNN* seperti *Fast R-CNN*, *Faster R-CNN*, *YOLO*, *AlexNet*, dan *VGGNet* telah meningkatkan akurasi dan kemampuan pemrosesan *real-time*. Dalam penelitian ini, mengevaluasi deteksi dan segmentasi manusia pada kumpulan data *MASK MADS* dan melaporkan metrik kinerja seperti *Average Precision (AP)* untuk segmentasi dan pelacakan manusia. Penelitian ini menyoroti pentingnya deteksi manusia yang akurat dalam video dan membahas tantangan seperti latar belakang yang kompleks, halangan, dan blur gerakan.

Human In Point Clouds With Synthetic Data

Penelitian Takmaz et al. (2022) berjudul “*3D Segmentation Of Humans In Point Clouds With Synthetic Data*” membahas tentang masalah utama dalam merekam dan memberi label pada manusia asli dalam lingkungan *indoor* yang kompleks memiliki dengan skala yang terbatas. Sebagai alternatif, pendekatan mensintesis data pelatihan dapat menjadi pilihan (Dai et al. 2017). Data pelatihan sintesis dilengkapi dengan label yang sempurna dan lebih mudah untuk mengendalikan variasi dan keragamannya. Pemindaian kedalaman manusia dalam skenario 3D juga lebih mudah dilakukan daripada merender manusia secara fotorealistik dalam gambar berwarna (Wood et al. 2021).

Dengan mengusulkan metode model berbasis transformer baru yang disebut *Human3D*, yang melakukan banyak tugas segmentasi manusia dalam 3D secara terintegrasi. Secara khusus, model yang secara langsung menangani segmentasi banyak bagian tubuh manusia dalam *point cloud* lingkungan yang realistis (Schult et al. 2023). *Human3D* didasarkan pada mekanisme baru yang menggunakan kueri dua langkah untuk mengelompokkan masks secara bersamaan berdasarkan ekspresi manusia dan bagian tubuh yang sesuai.

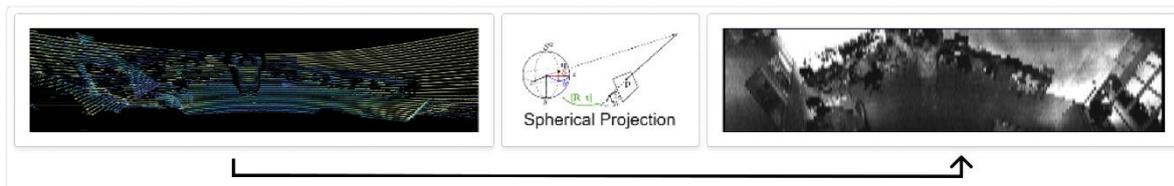
Hasil dari penelitian ini mengimplementasikan *Human3D*, model *end-to-end* yang melakukan segmentasi *3D real-time* dari beberapa bagian tubuh manusia dalam *point cloud* internal yang tidak beraturan. Tujuan utama adalah penggunaan kueri dua tingkat yang mewakili kasus manusia dan bagian tubuh dan menunjukkan bahwa dengan pendekatan yang relatif sederhana untuk menghasilkan data pelatihan sintetik secara signifikan meningkatkan kinerja segmentasi manusia 3D diberbagai tugas dan model.

Representing Full Human Data

Penelitian Krawczyk and Sitnik (2023) berjudul “*Segmentation of 3D Point Cloud Data Representing Full Human Body Geometry: A Review*” berfokus pada teknik segmentasi data tubuh manusia 3D dengan menggunakan berbagai metode dan algoritma. Masalahnya terletak pada tantangan dalam melakukan segmentasi model tubuh manusia secara akurat dari data *3D point cloud*. Segmentasi data *3D point cloud* yang mewakili tubuh manusia mencakup kebutuhan akan algoritma segmentasi *real-time* yang akurat (Plagemann and Koller 2010), pentingnya menjaga *edge* antar segmentasi, meningkatkan kinerja dalam posisi tubuh yang diuji atau ditingkatkan untuk bekerja secara efektif bahkan ketika tubuh dalam posisi tidak standar atau abnormal. Relevan dalam banyak aplikasi, seperti pencitraan medis yang dapat melakukan diagnosis dan pemeriksaan organ yang lebih akurat. Selain itu, *point cloud* telah terbukti berguna dalam memvisualisasikan pembedahan (Chen and Wang 2017) atau pelacakan gerak dimana tubuh subjek tidak selalu dalam posisi yang benar.

Dengan metode pendekatan *machine learning* untuk segmentasi data tubuh manusia 3D. Melibatkan tinjauan dan analisis komprehensif dari algoritma segmentasi yang ada, metode *machine learning*, dan teknik pengumpulan data, dengan tujuan untuk mengidentifikasi tantangan utama dan solusi potensial untuk mensegmentasi data *3D point cloud* yang mewakili tubuh manusia. Metode ini bertujuan untuk meningkatkan akurasi dan efisiensi segmentasi model tubuh manusia dari data *3D point cloud*.

Hasil dari penelitian ini menyajikan hasil penerapan algoritma segmentasi pada data *3D*



Gambar 1. Hasil Data Point Cloud Setelah Melalui Spherical Projection

point cloud yang merepresentasikan tubuh manusia. Studi ini membandingkan akurasi dan waktu komputasi berbagai algoritma, menyoroti keefektifan teknik *machine learning*, dan menyajikan visualisasi berbagai sistem pengumpulan data. Temuan penelitian ini fokus pada segmentasi data tubuh manusia 3D menggunakan pendekatan *machine learning*. Hasil eksperimen menunjukkan peningkatan akurasi dan efisiensi dalam mensegmentasi model tubuh manusia dari data *3D point cloud*. Algoritma segmentasi terbukti memberikan hasil yang lebih baik dibandingkan metode sebelumnya, menunjukkan akurasi dan konsistensi segmentasi yang lebih tinggi.

METODE

Untuk penelitian ini, peneliti membuat tahapan dan rancangan penelitian yang akan digunakan untuk membangun model segmentasi tersebut:

Tahap Pengumpulan Data

Dataset yang digunakan pada penelitian ini merupakan hasil scan *LiDAR* berupa *point cloud 3D*. Kumpulan data ini berisi 21 file, masing-masing dengan dua format file: *.JSON* dan *.PCAP*. Setiap file berisi *pose* manusia dengan *pose* berbeda-beda. Di antara 21 file tersebut, terdapat file dengan posisi dan *pose* yang sama, namun berbeda sudut depan, belakang, kiri, dan kanan. Detail file ditunjukkan di bawah ini. 4 file untuk *pose* berbaring, 1 file untuk *pose* berdiri, 4 file untuk *pose* berjongkok, 4 file untuk *pose* duduk, 4 file untuk *pose* berjongkok, 4 file untuk *pose* tangan.

Tahap Spherical Projection

Langkah pertama dalam mensegmentasi data 3D menggunakan *Spherical Projection* melibatkan konversi data menjadi gambar 2D, di mana setiap titik 3D diproyeksikan ke bidang 2D. Proses ini mensyaratkan agar data (1) dapat direpresentasikan dalam format yang mudah dikelola dan dapat diolah lebih lanjut dengan

menggunakan berbagai teknik pengolahan citra, (2) dapat menggunakan algoritma dan metode pencitraan 2D, dan (3) dapat mudah diubah, Gambar 2D yang dihasilkan dapat diterapkan pada *computer vision*.

Gambar 1. menunjukkan proyeksi titik dunia tiga dimensi (3D) yang dijelaskan oleh model sensor gambar ke koordinat piksel dua dimensi (2D) (Kim and Park 2022) Saat mengolah data *point cloud*, menggunakan sistem koordinat bola adalah pilihan yang lebih baik untuk jenis kumpulan data ini karena titik data *point cloud* objek berada di permukaan objek dan tidak memiliki titik pusat. Pada proses *coordinates transformation*, akan dilakukan konversi global *cartesian coordinates* (x,y,z) menjadi *local spherical coordinates* (ϕ, θ, r). Dengan rumus:

$$\rho = \sqrt{x^2 + y^2 + z^2}$$

$$\tan \phi = \frac{y}{x}$$

$$\cos \theta = \frac{\sqrt{x^2 + y^2}}{z}$$

Equation 1. Rumus *Coordinates Transformation*

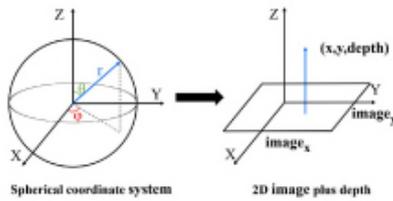
Nilai r_i adalah jarak atau radius dari titik tersebut ke titik pusat. Nilai θ_i dan ϕ_i masing-masing mewakili sudut horizontal dan sudut vertikal suatu titik dalam *spherical coordinates*. Setelah itu, terdapat proses *range images projection*. Pada proses *range images projection*, *spherical coordinates* (ϕ, θ, r) akan dikonversi menjadi *2D image plus depth*. Dapat dilihat pada Equation 2. untuk rincian Rumus *Range Image* dan Gambar 2. untuk rincian lebih jelas tentang Proses *Range Image*.

$$x_i = \text{round}\left(\frac{\phi_i}{2\pi} \times \text{image}_x\right)$$

$$y_i = \text{round}\left(\frac{\theta_i}{\pi} \times \text{image}_y\right)$$

$$\text{depth}_i = \text{round}(\tau_i)$$

Equation 2. Rumus *Range Image*



Gambar 2. Proses Range Image

Data mentah berupa *3D point cloud* dalam format file *.PCAP* dan *.JSON* sulit untuk diproses. Maka *spherical projection* ada untuk memudahkan proses tersebut. Dalam proses *spherical projection*, *3D Point Cloud* diubah menjadi *range image*.



Gambar 3. Proses Spherical Projection

Gambar 3. menunjukkan bagaimana *range image* dibuat menggunakan perangkat lunak *Ouster Studio* versi 2.0.2, yang dapat membaca file *3D point cloud* dengan format *.PCAP* dan *.JSON*. Data dikumpulkan dari seorang pria yang mengangkat tangannya dan memutarkannya 360 derajat selama 9 detik.

Selama durasi 9 detik, data yang didapat menjadi 71 gambar. Setelah itu data gambar dipotong agar lebih fokus kepada subjek, yaitu seorang pria yang berputar dengan kedua tangan diangkat. Gambar dipotong menjadi 126x32, lalu diberi label secara manual.

Pembagian data pada penelitian ini disajikan pada Tabel 1.

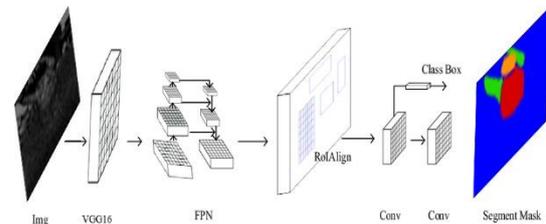
Tabel 1. Pembagian Data

No	Nama data	Data Gambar	Label Gambar	Total
1.	Data Train	71	71	142
2.	Data Test	71	71	142
3.	Data Val	71	71	142

Tahap Segmentasi Mask R-CNN

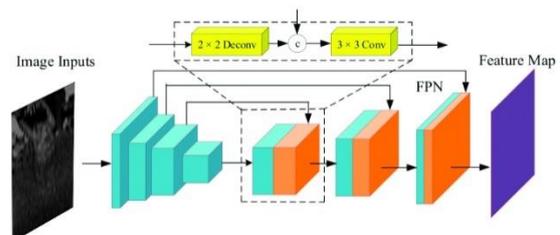
Jaringan *Mask R-CNN* terdiri dari dua tahap Tahap pertama adalah *Region Proposal Network (RPN)*, yang memprediksi kotak pembatas proposal objek berdasarkan anchor

box dengan skala yang berbeda dan ratio yang berbeda (K. Lin et al. 2020) Tahap kedua adalah detektor *R-CNN* yang menyempurnakan proposal-proposal ini, mengklasifikasikannya, dan menghitung segmentasi piksel pada tingkat proposal-proposal tersebut (He et al. 2017)



Gambar 4. Proses Model Mask R-CNN

Komponen juga menggunakan yaitu *encoder-decoder* dengan arsitektur *VGG16*, Model *CNN* yang dikembangkan menggunakan input multiskala yang berisi beberapa tingkat lapisan konvolusional. Lapisan ini dibagi menjadi dua blok: *encoder* dan *decoder*. Blok *encoder* menggabungkan dan menurunkan resolusi peta fitur yang berasal dari berbagai tingkatan lapisan konvolusional. Demikian pula, peta fitur yang diekstraksi oleh blok *decoder* digabungkan dan ditingkatkan resolusinya. Peta fitur terakhir digunakan untuk membedakan apakah suatu piksel palsu atau tidak. Dua kumpulan data yang tersedia digunakan untuk memvalidasi model (Kadam, Kotecha, and Ahirrao 2023)



Gambar 5. Arsitektur Encoder-Decoder

Tahap Evaluasi Model

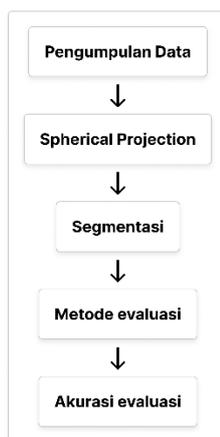
Pada tahap evaluasi, metrik evaluasi yang digunakan dengan metrik untuk model segmentasi, seperti *precision*, *recall* dan *F1-score*. Metrik-metrik ini memberikan gambaran tentang seberapa baik model dapat mengidentifikasi segmentasi objek pada gambar.

Tahap Evaluasi Akurasi

Akurasi merupakan metrik evaluasi yang mengukur sejauh mana model mampu

melakukan prediksi dengan benar. Selain akurasi juga menganalisis nilai *loss* pada model. *Loss* menggambarkan seberapa baik atau buruk model dalam membuat prediksi. Nilai *loss* yang rendah menunjukkan kinerja yang baik, sementara nilai *loss* tinggi menunjukkan kinerja yang buruk, Analisis akurasi dan *loss* dapat memberikan gambaran tentang efektivitas model. Dengan melakukan akurasi, *loss* dan hasil segmentasi, bias melihat tentang kinerja model.

Penelitian ini akan melewati beberapa tahapan. Antara lain: Pengumpulan data, *Spherical Projection*, Segmentasi, Metode evaluasi, dan Akurasi evaluasi. Tahapan-tahapan tersebut dapat dilihat pada Gambar 6.

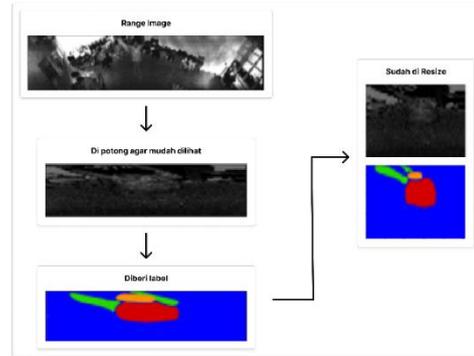


Gambar 6. Rancangan Penelitian Yang Diajukan

HASIL DAN PEMBAHASAN

Tahap Image Processing

Sebelum proses *spherical projection* dan proses segmentasi, *range image* harus dipotong. Setelah *range image* dipotong ke ukuran yang dapat dilihat, beri label secara manual menggunakan aplikasi desain. Agar dapat memasuki model yang digunakan, dalam *Mask R-CNN* ukuran gambar harus berukuran 192x256 piksel. Maka dari itu, data asli dan data yang telah diberi label akan di *resize* menjadi ukuran yang butuhkan agar dapat memberikan kinerja yang maksimal untuk pengujian dan validasi. Tahap *Image Processing* dijelaskan pada Gambar 7. yang menjelaskan secara rinci proses *Image processing*.



Gambar 7. Tahap *Image Processing*

Tahap Processing

Pada tahap ini, proses pelatihan model dilakukan menggunakan Kaggle menggunakan GPU P100 yang disediakan. Pada penelitian ini dilakukan pelatihan model *Mask R-CNN* untuk segmentasi bagian tubuh manusia untuk menentukan kepala, tangan dan badan dari manusia. Pembagian parameter tahap processing ini disajikan pada tabel 2.

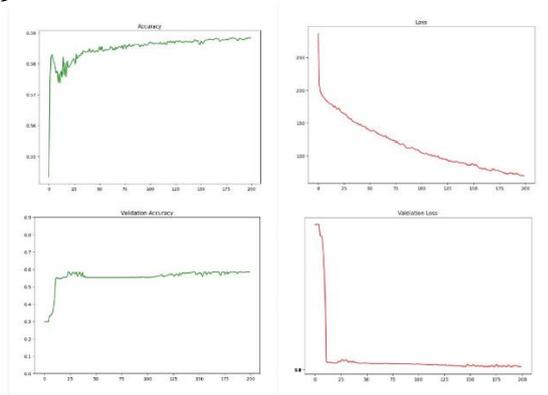
Tabel 2. Pembagian parameter

Parameter	Value
Convolutional Layer	27
Pooling Layer	5
Upsampling Layer	5
Dropout	0.5
Activation Function	Relu, Softmax
Optimizer	Adam
Loss Function	Categorical Crossentropy
Epoch	200
Batch Size	27
Learning Rate	0.0001

Proses segmentasi dilakukan menggunakan model *Mask R-CNN* dengan arsitektur *VGG16* dengan menggunakan parameter yang dijelaskan di Tabel 2. Pada proses *Encoder* pada model yang dibuat menggunakan 13 blok dan proses *Decoder* juga menggunakan 14 blok yang terdiri dari *ReLU Layer*, *Batch Normalization Layer*, *Dropout*, *Convolution Layer*, dan juga menggunakan 5 blok *Pooling Layer* dan 5 blok *Upsampling Layer* dengan filter kernel 3x3. *Optimizer* yang digunakan adalah adam dan loss function adalah *Categorical Crossentropy* dengan jumlah *batch* 18. Dengan proses pelatihan sebanyak 200 *Epochs*.

Tahap Uji Coba Dan Validasi

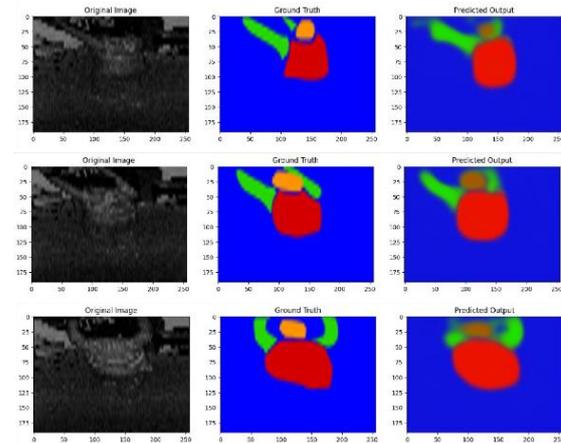
Dengan akurasi pengujian 59%, model ini cukup baik. Adapun akurasi dan *loss* yang ditentukan dari hasil pengujian dapat dilihat pada Gambar 8.



Gambar 8. Grafik Akurasi dan *Loss*

Akurasi pada gambar 8. merupakan metrik evaluasi yang mengukur sejauh mana model mampu melakukan prediksi. Dalam penelitian ini, akurasi diukur sebagai persentase dari jumlah prediksi yang benar dibandingkan dengan total jumlah prediksi pada data pengujian. Hasil akurasi mencapai 59%. Selain akurasi, menganalisis nilai *loss* pada model. *Loss* menggambarkan seberapa baik atau buruk model dalam membuat prediksi. Nilai *loss* yang rendah menunjukkan kinerja yang baik, sementara nilai *loss* tinggi menunjukkan kinerja yang buruk. Analisis akurasi dan *loss* dapat memberikan gambaran tentang efektivitas model.

Model *Mask R-CNN* yang sudah dirancang sesuai keperluan dalam penelitian ini kemudian dilakukan prediksi pada gambar secara acak. Adapun hasil prediksi gambar bisa dilihat di Gambar 9.



Gambar 9. Hasil Segmentasi dan Prediksi Citra

Dari Gambar 9. terlihat bahwa hasil segmentasi menggunakan model *Mask R-CNN* untuk mensegmentasi bagian tubuh manusia mencapai tujuan yang diharapkan yaitu mampu membedakan kepala, tangan, dan tubuh manusia. Dari segi hasil segmentasi terlihat masih terdapat beberapa kegagalan dan hal-hal yang tidak sesuai, seperti hasil segmentasi yang kurang tepat dan prediksi yang kurang sesuai, seperti kepala terlalu kecil dan badan yang terlalu besar.

SIMPULAN DAN SARAN

Pada penelitian ini, dapat dirangkum beberapa kesimpulan yang dapat diambil, sebagai berikut :

1. Hasil pelatihan model menunjukkan bahwa penggunaan model *Mask R-CNN* dengan arsitektur *VGG16* untuk segmentasi bagian tubuh manusia menggunakan data *3D Point Cloud* memberikan hasil yang baik mengingat akurasi yang ditampilkan dengan nilai akurasi sebesar 59%
2. Memperbaiki model dengan mengevaluasi pilihan arsitektur selain *Mask R-CNN* guna untuk menentukan apakah terdapat peningkatan yang berarti dalam hasil segmentasi.
3. Mengidentifikasi dan mengatasi penyebab hasil segmentasi yang kurang sesuai dengan menyesuaikan algoritma atau proses pelatihan guna diperlukan untuk mencapai hasil yang lebih jelas.

DAFTAR PUSTAKA

- Bellynza, Ken Ayu, and Hadi Syaputra. 2022. "Objek Deteksi Burung Lovebird Menggunakan Instance Segmentation Mask

- R-Cnn.” *Bina Darma Conference on Computer Science(BDCCS2022)* 4 (1).
- Chen, Guo Dong, and Fei Fei Wang. 2017. “Medical Data Point Clouds Reconstruction Algorithm Based on Tensor Product B-Spline Approximation in Virtual Surgery.” *Journal of Medical and Biological Engineering* 37 (2). <https://doi.org/10.1007/s40846-016-0211-3>.
- Dai, Angela, Angel X Chang, Manolis Savva, Maciej Halber, and Thomas Funkhouser. 2017. “ScanNet : Richly-Annotated 3D Reconstructions of Indoor Scenes.” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, no. 10.1109/CVPR.2017.261.
- Danielczuk, Michael, Matthew Matl, Saurabh Gupta, Andrew Li, Andrew Lee, Jeffrey Mahler, and Ken Goldberg. 2019. “Segmenting Unknown 3D Objects from Real Depth Images Using Mask R-CNN Trained on Synthetic Data.” *Proceedings - IEEE International Conference on Robotics and Automation* 2019-May: 7283–90. <https://doi.org/10.1109/ICRA.2019.8793744>.
- He, Kaiming, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. “Mask R-CNN.”
- Hochman, J., H. R. Bourne, P. Coffino, P. A. Insel, L. Krasny, and K. L. Melmon. 1977. “Subunit Interaction in Cyclic AMP Dependent Protein Kinase of Mutant Lymphoma Cells.” *Proceedings of the National Academy of Sciences of the United States of America* 74 (3). <https://doi.org/10.1073/pnas.74.3.1167>.
- Hong Hai, Hoang, and Tran Bao Long. 2021. “Improve of Mask R-CNN in Edge Segmentation.” *JST: Engineering and Technology for Sustainable Development* 31 (3): 97–104. <https://doi.org/10.51316/jst.151.etsd.2021.31.3.17>.
- Kadam, Kalyani Dhananjay, Ketan Kotecha, and Swati Ahirrao. 2023. “Retracted: Efficient Approach towards Detection and Identification of Copy Move and Image Splicing Forgeries Using Mask R-CNN with MobileNet V1.” *Computational Intelligence and Neuroscience* 9786963 (13). <https://doi.org/10.1155/2023/9786963>.
- Kim, Songeun, and Soon Yong Park. 2022. “Expandable Spherical Projection and Feature Concatenation Methods for Real-Time Road Object Detection Using Fisheye Image.” *Applied Sciences (Switzerland)* 12 (5). <https://doi.org/10.3390/app12052403>.
- Krähenbühl, Philipp, and Vladlen Koltun. 2014. “Geodesic Object Proposals.” *European Conference on Computer Vision (ECCV)* 8693. https://doi.org/10.1007/978-3-319-10602-1_47.
- Krawczyk, Damian, and Robert Sitnik. 2023. “Segmentation of 3D Point Cloud Data Representing Full Human Body Geometry: A Review.” *Pattern Recognition* 139: 109444. <https://doi.org/10.1016/j.patcog.2023.109444>.
- Le, Van Hung, and Rafal Scherer. 2021. “Human Segmentation and Tracking Survey on Masks for Mads Dataset.” *Sensors* 21 (24): 1–22. <https://doi.org/10.3390/s21248397>.
- Lin, Kailian, Huimin Zhao, Jujian Lü, Jin Zhan, Xiaoyong Liu, and Rongjun Chen. 2020. “Face Detection and Segmentation Method Based on Mask R-CNN.” *Jisuanji Gongcheng/Computer Engineering* 46 (6). <https://doi.org/10.19678/j.issn.1000-3428.0054566>.
- Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. “Microsoft COCO: Common Objects in Context.” *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland* 8693 (10.1007/978-3-319-10602-1_48).
- Liu, Maohua, Yue Shao, Ruren Li, Yan Wang, Xiubo Sun, Jingkuan Wang, and Yingchun You. 2020. “Method for Extraction of Airborne LiDAR Point Cloud Buildings Based on Segmentation.” *PLoS ONE* 15 (5). <https://doi.org/10.1371/journal.pone.0232778>.
- Marpaung, Faridawaty, Arnita, Fitrihuda Aulia, Nita Suryani, and Rinjani Cyra Nabila. 2022. *COMPUTER VISION DAN PENGOLAHAN CITRA DIGITAL*.
- Plagemann, Christian, and Daphne Koller. 2010. “Real-Time Identification and Localization of Body Parts from Depth Images.” *IEEE International Conference on Robotics and Automation, Anchorage, AK, USA* 3108–3 (10.1109/ROBOT.2010.5509559).
- Rao, Deepak, Quoc V Le, Thanathorn Phoka, Morgan Quigley, Attawith Sudsang, and Andrew Y Ng. 2010. “Grasping Novel Objects with Depth Segmentation.” *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, Taipei, Taiwan* 2578 (10.1109/IROS.2010.5650493).
- Ryselis, Karolis, Tomas Blažauskas, Robertas Damaševičius, and Rytis Maskeliūnas. 2022. “Computer-Aided Depth Video Stream Masking Framework for Human Body Segmentation in Depth Sensor Images.” *Sensors* 22 (9). <https://doi.org/10.3390/s22093531>.
- Schult, Jonas, Francis Engelmann, Alexander Hermans, Or Litany, Siyu Tang, and Bastian Leibe. 2023. “Mask3D: Mask Transformer for 3D Semantic Instance Segmentation.”

- Proceedings - IEEE International Conference on Robotics and Automation 2023.*
<https://doi.org/10.1109/ICRA48891.2023.10160590>.
- Takmaz, Ayça, Jonas Schult, Irem Kaftan, Mertcan Akçay, Bastian Leibe, Robert Sumner, Francis Engelmann, and Siyu Tang. 2022. "3D Segmentation of Humans in Point Clouds with Synthetic Data."
<http://arxiv.org/abs/2212.00786>.
- Uckermann, Andre, Christof Elbrechter, Robert Haschke, and Helge Ritter. 2012. "3D Scene Segmentation for Autonomous Robot Grasping." *IEEE International Conference on Intelligent Robots and Systems*, no. October 2012.
<https://doi.org/10.1109/IROS.2012.6385692>.
- Wood, Erroll, Tadas Baltruš, Hewitt Sebastian, Matthew Johnson, Virginia Estellers, Thomas J Cashman, and Jamie Shotton. 2021. "Fake It till You Make It : Face Analysis in the Wild Using Synthetic Data Alone." *IEEE/CVF International Conference on Computer Vision (ICCV)*, no. 10.1109/ICCV48922.2021.00366.
- Wu, Xin, Shiguang Wen, and Yuan ai Xie. 2019. *Improvement of Mask-RCNN Object Segmentation Algorithm. Intelligent Robotics and Applications. ICIRA 2019*. Vol. 11740. Springer International Publishing.
https://doi.org/10.1007/978-3-030-27526-6_51.
- Xu, Jiaxin, Rui Wang, Vaibhav Rakheja, and Computing Science. 2019. "Literature Review: Human Segmentation with Static Camera." *Computer Vision and Pattern Recognition 1*
(<https://doi.org/10.48550/arXiv.1910.12945>).
- Yao, R U I, Guosheng Lin, Shixiong Xia, Jiaqi Zhao, and Yong Zhou. 2019. "Video Object Segmentation and Tracking : A Survey." *ACM Transactions on Intelligent Systems and Technology 11*
(<https://doi.org/10.1145/3391743>).
- Zhang, Song-hai, Ruilong Li, Xin Dong, Paul Rosin, Zixi Cai, Xi Han, Dingcheng Yang, Haozhi Huang, and Shi-min Hu. 2019. "Pose2Seg : Detection Free Human Instance Segmentation." *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 898 (10.1109/CVPR.2019.00098).